Journal of the Brazilian Computer Society, 2026, 32:1, doi: 10.5753/jbcs.2026.4471

Proposing and Evaluating a MALTU Extension for UbiComp and IoT Systems

Carlos E. Almeida Feitosa [**Federal University of Ceará** | ***eduardoalmeida8246@gmail.com***]
Francisco Das C. Moreira Catunda Neto [**Federal University of Ceará** | ***fcmcneto@gmail.com***]
Ingrid Teixeira Monteiro [**Federal University of Ceará** | ***ingrid@ufc.br***]

Quixadá Campus, Federal University of Ceará, Av. José de Freitas Queiroz, 5003, Cedro, Quixadá, CE, 63902-580, Brazil.

Received: 29 April 2024 • **Accepted:** 05 May 2025 • **Published:** 31 July 2026

Abstract Systems developed following the Ubiquitous Computing (UbiComp) and Internet of Things (IoT) paradigms have certain specific quality-of-use characteristics that traditional Human-Computer Interaction (HCI) evaluation methods may not be able to address, such as context-awareness, invisibility, attentiveness, calmness, mobility and synchronicity, since most HCI methods were traditionally created to evaluate desktop or web systems. The Model for Evaluating Interaction in Social Systems Based on the User's Textual Language (MALTU) evaluates usability and user experience based on comments left by users on social systems (e.g. Amazon, Google Play Store, Facebook forums). This paper presents an extension in the model to address these UbiComp/IoT quality characteristics, describing the entire extension creation process and presenting the results of the first evaluations carried out using the extended MALTU. Before proposing the extension, three evaluations were performed with the original MALTU. Then, three evaluators performed two evaluations each on the same smart device: the first one was the application of the extended MALTU, and the second one was a heuristic evaluation using a specific set of heuristics for UbiComp systems. After that, a focus group was held in order to verify if the extended MALTU was capable of identifying typical problems of ubiquitous devices considering its quality characteristics. With the first version of the extension completed and published, a round of external evaluations was carried out with the extended MALTU, by authors not participating in the design of the extension. The results indicate that the extension was able to identify interaction problems in ubiquitous devices, such as Context Awareness, Synchronicity and Mobility, even with difficulties in understanding the concepts of quality characteristics of ubiquitous and IoT systems.

Keywords: Ubiquitous systems, IoT, HCI, Evaluation, MALTU

1 Introduction

The Ubiquitous Computing (UbiComp) proposes systems composed of interconnected smart devices that are available in a nonintrusive and continuous manner in the user's physical environment, aiming for technology to remain invisible and cause minimal possible distractions, capable of monitoring behaviors and providing services without realizing that we are interacting with a computer [Andrade *et al.*, 2017; Carvalho, 2017; de Souza Filho *et al.*, 2022].

One of the main characteristics of UbiComp/IoT systems is that they depart from the desktop environment to be incorporated into everyday objects. They have specific capabilities, such as initiating interaction with the user and being present in various internet-connected objects [Carvalho, 2019]. Such characteristics should be considered to support the user in their daily tasks [Andrade *et al.*, 2017], and as a result, new measures must be applied for a comprehensive evaluation of Human-Computer Interaction (HCI) quality in these applications [Carvalho *et al.*, 2018; Santos, 2014].

In this context, the HCI field studies ways to develop and evaluate interactive systems considering various quality of use criteria such as usability, accessibility and user experience [Barbosa and Silva, 2010; de Souza Filho *et al.*, 2020]. However, traditional HCI evaluations may not be suitable for assessing UbiComp and IoT systems, as HCI research began in the desktop computing world, where users had only one computer to accomplish their goals, but with these new paradigms, users started using multiple devices and interfaces: mobile phones, tablets, laptops, smartwatches [Carvalho, 2019].

Another important point regarding the evaluation of these systems is that certain specific characteristics are difficult to assess because it requires a lot of time and experience to identify problems [de Souza Filho *et al.*, 2020]. An example is context sensitivity, which refers to the system's ability to perceive contextual information from the user and the environment. To identify problems related to this quality characteristic, it is necessary to use the system in as many contexts as possible.

In the literature, it is possible to find frameworks [Karvonen and Kujala, 2014; Kourouthanassis *et al.*, 2008; Metsis *et al.*, 2008; Ntoa *et al.*, 2021], quality models [Carvalho *et al.*, 2017], and specific heuristics [Rocha, 2017] for evaluating ubiquitous systems. However, none of the works consider evaluation from the user's textual language perspective, considering that in recent years it has become common to use websites and applications for product evaluation and complaints [Mendes, 2015], such as Amazon, Booking, Decolar, and Reclameaqui. Furthermore, there are works that focus on user narratives for usability and user experience evaluation

[Korhonen *et al*., 2010; Mendes, 2015; Olsson and Salo, 2012; Tuch *et al*., 2013], but there are no models that consider user textual narratives specifically for evaluating the interaction of UbiComp and IoT systems.

Thus, this article presents the process of creation of an extension of the MALTU model (Model for Evaluation of Interaction in Social Systems from User Textual Language) that evaluates interaction based on user comments left on websites, applications, or forums addressing system usage [Mendes, 2015]. The model works by classifying these comments into defined categories and can be extended by adding a step or category to the original method. This addition can deepen the analysis and address other aspects of interaction, such as Attention, Mobility, Invisibility, Context-Awareness, Calmness, and Synchronicity, which is a a set of quality use characteristics for UbiComp and IoT systems know as AMICCaS [Carvalho, 2019].

In this article, the following research questions are addressed:

- **RQ1**. Is it possible to satisfactorily extend the MALTU model to meet the UbiComp/IoT system quality characteristics (AMICCaS)?
- **RQ2**. How does the proposed extension of MALTU help HCI professionals reflect on the quality use characteristics of UbiComp/IoT systems?

To answer RQ1, a set of three evaluations of IoT systems were carried out, using the original MALTU, in an attempt to identify user comments related to AMICCaS. After this study, the MALTU extension was developed.

To answer RQ2, two evaluations of the extended model were carried out: 1) application of the extended MALTU compared to an evaluation with specific IoT heuristics (three evaluators evaluating the same device) for the purpose of validating the extension; 2) independent application of the extended MALTU by an evaluator external to the group (one evaluator evaluating two different devices).

In the first study, three invited evaluators participated. It was divided into three parts: in the first phase, the participants individually evaluated the same device and the same set of user comments separately, without contact with each other's responses. In the second part, the evaluators applied a heuristic evaluation with heuristics for ubiquitous systems on the same device evaluated with the extended MALTU. Finally, in the third part, a focus group was conducted to assess the evaluators' opinions on the two evaluation models.

The second study delves into two ubiquitous systems aimed at simplifying blood glucose monitoring: the G-TECH APP and the FreeStyle Libre. Through an in-depth examination of their context of use, system domain, and evaluation objectives, the research aims to assess their effectiveness and user experience. The G-TECH APP serves individuals seeking convenient blood glucose monitoring, integrating with a Bluetooth glucometer for on-the-go readings, primarily targeting those with type 1 diabetes. In contrast, FreeStyle Libre offers a more advanced approach with a subcutaneous sensor for continuous readings, providing real-time insights into glucose levels. Both systems target individuals with type 1 diabetes, offering improved functionalities like continuous readings and third-party data access via the app.

This paper is an extended version of [omitted by anonymous review] published in IHC 2023 (XXII Brazilian Symposium on Human Factors in Computing Systems). This article is organized as follows: Sections 2, 3, and 4 present the main concepts that underpinned the work. Sections 5 and 6 describe the methodological procedures and the results, respectively. Section 7 presents the initial uses of the extended MALTU. Section 8 presents the final considerations and future work.

2 Quality in UbiComp and IoT systems

UbiComp (Ubiquitous Computing) and IoT (Internet of Things) systems consist of interconnected objects that are also connected to the Internet and possess capabilities to detect and act upon context, distinguishing themselves from traditional desktop and web systems in terms of quality characteristics and interaction [Carvalho, 2019; de Souza Filho *et al*., 2022]. In these paradigms, interaction may occur through computers present in various forms, sizes, interfaces, and everyday objects such as washing machines, lamps, watches, locks, among others, with the main motivation being to assist the user in their daily activities without them realizing they are dealing with a computational infrastructure [Carvalho, 2019; Andrade *et al*., 2017; Siewerdt *et al*., 2016].

Thus, the technology completely changes the way users interact with it, as it is necessary for the system to be invisible or cause minimal distractions, in addition to being able to monitor the context and the user themselves to provide the appropriate services, not to mention that the system itself can initiate interaction, thanks to its various detection capabilities [Carvalho, 2017, 2019].

Despite IoT inheriting many of ubiquitous computing's goals, such as presence in everyday objects and environments and the ability to perceive and act in the real world, it entails a broader concept that encompasses ubiquitous systems [Carvalho, 2019]. The terms ”connection” and ”internet” are almost universally included in definitions, suggesting a system comprised of numerous objects. Thus, IoT relates not only to daily objects capable of detecting and acting upon the context but also to an expanding array of objects interconnected through network connections [Carvalho, 2019].

In Carvalho [2019], a set of six characteristics considered essential in the quality of interaction for UbiComp and IoT systems, named AMICCAS, are presented: Attention, Mobility, Invisibility, Context-Awareness, Calmness, and Synchronicity.

Attention refers to the user's ability to focus more on their daily activities, such as walking and driving, than on technology. **Mobility** pertains to the continuous use of the system as the user moves through different devices and networks. **Invisibility** relates to the system's ability to hide itself so that the user is not aware of it, as interaction is carried out through natural interfaces. **Context-Awareness** is the system's capability to perceive contextual information related to the user and the system environment. **Synchronicity** refers to the ability of things to be synchronized with one another. Finally, **Calmness** is related to the ability to prevent users from feeling overwhelmed by the system.

These characteristics directly impact the quality of interaction in UbiComp and IoT applications, and therefore they must be considered in HCI evaluations of these applications [Andrade *et al*., 2017].

Moreover, the scope of [Carvalho, 2019] includes UbiComp and IoT systems; thus, the AMICCaS reflect interaction quality characteristics that lie at this intersection.

3 Evaluation of UbiComp and IoT systems

- Clarify whether the meetings for the first study were conducted online or in face to face, and whether they were sequential or had intervals.

The HCI field studies how to develop and evaluate interactive systems concerning various aspects of use quality, such as accessibility, usability, and user experience [Barbosa and Silva, 2010]. However, when addressing UbiComp and IoT systems, the field continues its initial advancements [Brito, 2018] regarding the types of methods available to assess all possible variables influencing the quality of interactions with these systems.

de Souza Filho *et al*. [2022], for instance, explained how the concept of invisibility has evolved historically from two perspectives: the attributes related to invisibility and the evaluative approaches that consider this attribute. Among the main findings, there is a catalog with 13 types of evaluations that account for invisibility in the evaluation process: six frameworks, two questionnaire instruments, two quality models, a set of measures, and two sets of heuristics.

Melo *et al*. [2022] investigated long-term UX evaluations in the context of IoT systems. They conducted a snowballing procedure and systematic literature mapping to understand whether and how long-term UX evaluations have been conducted in IoT and the challenges and opportunities in conducting longitudinal studies in this field. They also interviewed five software engineering professionals to learn about their IoT user-testing and evaluation perspectives. As part of the results, the authors identified three challenges on longitudinal IoT studies: 1) Application of privacy-oriented methods, 2) Conducting Evaluations in real Contexts, and 3) Choice of Appropriate Constructs

In Rocha [2017], the development of specific heuristics for UbiComp systems, based on Nielsen's heuristics and termed HUbis: Usability Heuristics for Ubiquitous Systems, is presented. Further, de Souza Filho *et al*. [2020] compares heuristic evaluation using the heuristics developed by Rocha [2017] with the MALTU model [Mendes, 2015], to determine the suitability of these two methods for evaluating usability and user experience in UbiComp systems. Among the main contributions, the authors present a set of recommendations for evaluating ubiquitous systems, along with a description of the differences and similarities observed in the two evaluations.

Brito [2018] explores the MALTU model in comparison to the heuristic evaluation from [Rocha, 2017] also to assess usability and user experience in UbiComp and IoT systems. The principal findings of this study indicate that the model efficiently and quickly evaluates usability and user experience issues, although certain problems were only addressed through heuristic evaluation. Additionally, the work discusses that the MALTU results were comprehensive, whereas those from the heuristic evaluation were very specific, similar to [de Souza Filho *et al*., 2020], pointing out not only the problems but also their specific locations and identifying issues only encountered during everyday system use.

Heuristic evaluation is an inspection method that guides evaluators to examine the interface to identify usability issues [Barbosa and Silva, 2010]. It is guided by a set of heuristics and comprises the following stages: preparation, data collection, interpretation, consolidation of results, and reporting.

However, it is important to note that the original heuristic evaluation method was proposed for assessing traditional systems and interfaces [Nielsen and Molich, 1990], and thus, certain characteristics of other application domains may not be adequately captured [de Souza Filho *et al*., 2020; Quiñones and Rusu, 2017]. To assess the usability of UbiComp/IoT systems, it is necessary to consider aspects common to all computational systems but also their particular characteristics [Rocha, 2017].

Thus, [Rocha, 2017] proposes, based on Nielsen's heuristics, a heuristic evaluation specifically tailored for ubiquitous systems with 15 unique heuristics (HUbis) (Table 1). The method consists of three stages, as described below:

- Preparation: In the preparation session, the facilitator should invite 3 to 5 evaluators and explain the portion of the system that will be assessed. The evaluators should have contact with the application, usage scenarios, and functionalities to be evaluated.
- Evaluation: The evaluators will be in contact with the device to become familiar with it and use it in all possible usage scenarios. After this recognition phase, the evaluators check the HUbis and observe if they are being followed.
- Results: The facilitator should schedule a meeting with all evaluators to discuss, prioritize the identified problems, and suggest solutions.

In the evaluation stage, a spreadsheet is provided containing the heuristics, their detailed descriptions, and three fields with the following options: "yes" when the application complies with the usability recommendations of the heuristic; "no" otherwise; and "n/a" if the application does not have aspects related to the investigated heuristic. Additionally, there is also a record of the problem in the same document, which includes severity, location where it occurred, and a detailed description.

For the evaluation of UbiComp and IoT systems Rocha [2017] developed specific heuristics that address the characteristics quality of these systems. All heuristics are described below. The excerpts of text in italics indicate the adaptations made by Rocha [2017] to the original Nielsen's heuristics to cover UbiComp and IoT systems. The last five heuristics are exclusive to this type of systems.

- **UH1: Visibility of system status**: The system should always provide feedback to the user in response to an interaction. This feedback should not interrupt the user in their current activity or overload them with information but should exist in the form of a perceptible change in some of the interface interaction modalities.

- **UH2: Match between the system and the real world**: The system should speak the user's language, with words, symbols, concepts, and interactions familiar to the user, rather than system-oriented. It should follow real-world conventions, making information seem logical and natural and easily achieve the intended goal.
- **UH3: User control and freedom**: The user should feel at ease to interact with the application or not. When the user wishes to interact with the system, they should be in control, able to interrupt a task or undo an operation at any time and return to the previous state. *When the application interacts with the user in a particular context, the user should not feel compelled to respond to the interaction and should have the option to ignore it to maintain focus on their current activity.* All these actions should be clearly marked in the system, and their visualization, if any, should maintain consistency throughout the application.
- **UH4: Consistency and standards**: Application interfaces, data inputs, ways of interacting, *or adapting to the context* should be consistent and follow conventions and standards familiar to the user, so that the software can be understood, learned, and used.
- **UH5: Error prevention**: It is important to know the situations that most cause errors and modify interfaces and interactions so that users do not make these errors when interacting with the application. *Additionally, the application must be able to maintain its services and performance always available when used by one or more users, in specific or adverse conditions.*
- **UH6: Recognition rather than recall**; Definition: When a dialogue or interaction is available, the system should minimize the user's memory load, leaving objects, actions, and options *available for at least one of the user's senses.*
- **UH7: Flexibility and efficiency of use**: The application should provide shortcuts to speed up interaction, in order to reduce the effort required to achieve the intended goal, especially for the advanced user. The integrity of functionality must be maintained when using shortcuts or not. Additionally, the system should be flexible, giving the user the possibility to customize settings according to their needs and experiences.
- **UH8: Aesthetic and minimalist design**: Dialogues should contain only relevant and necessary information, neither more nor less. Every extra unit of information in a dialogue competes with relevant units of information. *The sequence of interaction and access to components and functionalities should be available depending on the context, in a simple and natural way.*
- **UH9: Help users recognize, diagnose, and recover from errors**: Error messages should be simple and expressed in clear language (without codes), accurately indicating the problem and suggesting a solution constructively. *In addition to texts, error messages can be displayed in other formats available on mobile devices, such as images, audio, vibration. Error messages should guide the user cautiously, without stress, so that they do not stop using the system.*
- **UH10: Help and documentation**: Ideally, the application should be so easy to use (intuitive) that it does not need help or documentation. If necessary, help should be easily accessible, focused on the user's current activity. Help guidelines should be simple and objective.
- ***UH11: Mobility and devices***: Ubiquitous applications must maintain their functionality with the physical movement of the user and on devices with different capabilities. Aspects such as wireless network, device connection, screen size, limited hardware capacity, and power capacity are factors that the application needs to take into consideration to adjust during use without causing inconvenience to the user.
- ***UH12: Privacy and security***: The application must be able to keep saved and protected information so that there is no risk of damage in a specific usage context. Information must be transported and stored securely, as well as the application's access controls.
- ***UH13: Invisibility and transparency***: The system must be able to hide computational components so that users do not worry about them. Interactions should occur through natural interfaces.
- ***UH14: Context awareness and adaptive interfaces***: The ubiquitous application must react according to the context information the user encounters. Interfaces must adapt to these contexts and bring only relevant information in a way that facilitates the use of the system.
- ***UH15: Sensors and data input***: It should be verified whether data input, whether by the user or captured from the sensor, is effective and happening naturally for the user. The application must operate correctly in the presence of invalid inputs or stressful environmental conditions.

4 MALTU and the HCI evaluating from user comments

MALTU (Model for Assessing Interaction in Social Systems from User Textual Language) is a model that addresses an overall view of evaluation based on spontaneous texts related to system usage posted by users in Social Systems (SS), encompassing two usability quality criteria: user experience and usability [Mendes, 2015; Mendes and Furtado, 2018; de Souza Filho *et al.*, 2020]. Social Systems are the places where spontaneous texts associated with usage are reported by users [Mendes and Furtado, 2018]; in MALTU, these texts are called PRUs (Postings Related to Usage).

The model consists of five sequential steps: (1) Definition of the evaluation context; (2) Extraction of PRUs; (3) Classification of PRUs; (4) Interpretation of results; and (5) Reporting of results.

The definition of the context involves delimiting the context of use, describing the domain to be evaluated, and stating the evaluation objective. PRU extraction can be performed automatically by the UUX-posts tool [1] or manually through a database (e.g., CSV spreadsheet).

[1] `http://uuxposts.russas.ufc.br/`

For PRU classification, the model already provides a standardized set of categories. For instance, under "type" it can be divided into: criticism, praise, doubt, comparison, suggestion, or help; "intention" is divided into: visceral, behavioral, or reflexive; "functionality" indicates which function or feature the user is referring to in the PRUs; "usability" is divided into: satisfaction, safety, learning; and "user experience". Definitions for each category and subcategory can be found in [Mendes, 2015].

In interpreting the results, evaluators should pay attention to the frequency of each category, highlighting their relationships and identifying patterns. Graphs and tables can be generated for better visualization and understanding of the data. Finally, evaluators should produce a report describing the results obtained from the evaluation.

Several works mentioned in previous sections used the model to assess usability and user experience of both traditional and ubiquitous systems. In Freitas *et al*. [2016], the authors present a textual evaluation of Spotify using MALTU, and the results showed that the application receives many criticisms for various reasons, predominantly support and effectiveness issues leading to user frustration. Santos *et al*. [2024] discusses the analysis of user comments in an automatic translation application from Portuguese to Brazilian Sign Language (Libras). The analysis followed the MALTU methodology, evaluating interaction in social systems through users' textual language.

In Brito [2018]; de Souza Filho *et al*. [2020, 2018], the authors used the model in the ubiquitous context; however, none of them proposed a modification to the original MALTU to adapt it to encompass certain interaction characteristics present in the evaluated ubiquitous systems. The evaluations in the three works used more than one method to obtain diverse results, with one of these methods specific to the ubiquitous domain. For example, in [Brito, 2018; de Souza Filho *et al*., 2020], the authors applied MALTU and heuristic evaluation with HUbis, while in [de Souza Filho *et al*., 2018], in addition to the two previous methods, the authors also performed a usability test. Among the main findings, the works agree that the level of detail in MALTU regarding the identified problems is low, with more general descriptions. However, the model excels in speed and low cost compared to heuristic evaluation with HUbis. Additionally, in [de Souza Filho *et al*., 2018, 2020], the authors pointed out that MALTU is capable of assessing systems in the ubiquitous domain, reporting problems that also emerged in heuristic evaluation and need time to emerge.

In addition to the use of MALTU, there are many other HCI studies being carried out evaluating user comments, with different research objectives. Diniz *et al*. [2022] investigated the presence of usability heuristic issues through user reviews. More recently, researchers have been interested in investigating accessibility aspects in user comments. Santos *et al*. [2024] analyzed accessibility reviews of mobile apps to show evidence that updates may introduce barriers that make the app less accessible than its previous version according to users' perceptions. There are also research on accessibility for users with Autism Spectrum Disorder (ASD) [Santiago and Marques, 2022, 2023]. Considering the potential of user reviews, the authors presented a textual analysis of reviews extracted from eight educational applications available in Portuguese with a focus on autistic children.

5 Extensions for specific domains

The MALTU was grounded using the UUX (Usability and User Experience) quality criteria. However, it allows for extensions to other quality criteria or aspects of interest, such as collaboration, culture, communicability, among others. The model can still be evolved; its activities can be further detailed to facilitate its use aiming at improving its execution, or new ways of extracting or classifying PRUs, such as the use of rules based on sentence formation, can be proposed [Mendes, 2015].

The use of extensions of HCI methods is common to encompass peculiar characteristics of certain domains. In de Souza *et al*. [2015], for example, the authors extended MoLIC (Modeling Language for Interaction as Conversation), grounded by Semiotic Engineering [De Souza, 2005], with the intention of proposing a solution for collaborative interaction modeling design. The advent of collaborative systems brought new challenges to traditional HCI evaluation methods; thus, new models were proposed to represent collaboration.

Another work [Oliveira *et al*., 2008] also makes use of extensions to address specific characteristics. In this work, the authors proposed an extension of SIM (Semiotic Inspection Method) that considers how the designer's metamessage is being transmitted in the interface. The method was extended because the evaluated system is related to the educational environment. This means that in addition to the designer-user metamessage, it is necessary to also evaluate the teacher-student metamessage. The extension produced in the article allowed evaluating communicability from the teacher's point of view.

6 Extending MALTU to AMICCaS

In this section, we present the procedures followed to propose the MALTU extension. It is divided into two parts: the first describes the study and analysis of the original MALTU and the second discusses the creation of the extension.

6.1 Methodology

In the first step, three ubiquitous devices were selected for evaluation by the first author of the article using the original MALTU, based on the criteria outlined in [Brito, 2018], as described below:

- Have an associated mobile application published on the PlayStore.
- Have at least one discussion forum such as Facebook groups, comments on virtual stores, among others.
- Have user-posted comments available about their experiences.
- Belong to different categories or subcategories according to the following taxonomy [Asghari *et al*., 2019]: healthcare, environmental, smart city, commercial, industrial, and general aspects.

The first selected device was the Smart Bulb (Figure 1 from the "Smart Home" category, capable of turning on, off, and adjusting light intensity at established times, offering 16 million colors to customize the environment, according to the manufacturer, and with the Multilaser Liv application.

Figure 1. Smart Bulb

The second was the smart water sprinkler (Figure 2) from Rainpoint, which can be classified as both "smart home" and "smart agriculture", featuring automated irrigation functions allowing control over water quantity, scheduling, and planning irrigation, integrating with other devices, designed to operate in residences as well as in places requiring automated irrigation such as gardens, estates, and farms.

Figure 2. smart water sprinkler

The third device was the Smart Lock (Figure 3 from Locky, classified as "security and emergency", intended for residences but also applicable in offices, universities, and other establishments. It features a biometric sensor, digital keypad, access code generator for authorized individuals to enter the house, integration with other devices, and an application enabling remote door opening or locking.

The original MALTU was applied to the three devices, considering different definitions of the context of use for each system. PRUs were extracted from the period 2020 to 2022 and manually classified in a Google spreadsheet; the results served as input for the extension creation.

For the evaluation of the smart bulb, 88 PRUs with ratings of up to three stars were manually collected from the Amazon.com website between 2020 and 2022. During the classification, some PRUs could be categorized into more than one type. For instance, one user reported issues with

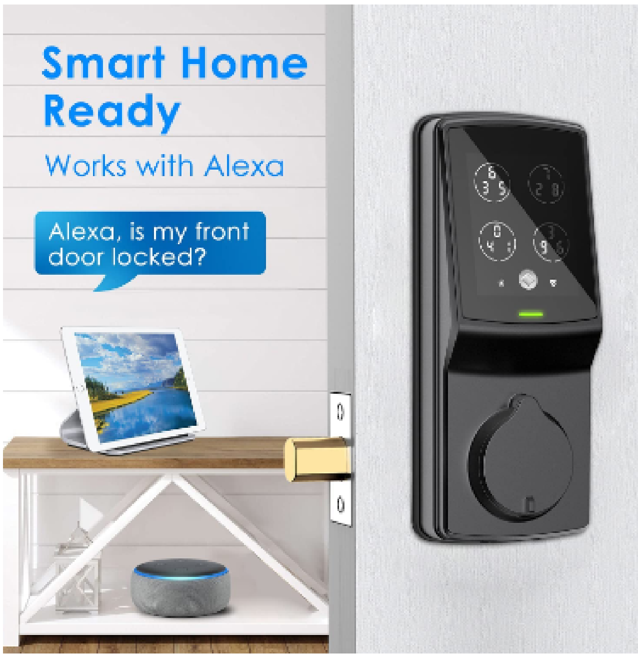

Figure 3. Smart Lock

the quality and instability of a product but also made positive comparisons with other brands. Users frequently mentioned problems with the bulbs' lighting, such as low durability and weak illumination, along with issues of internet connection and integration with devices like Alexa and Smart TVs. These problems were related to the Mobility characteristic of ubiquitous systems, which refers to the continuous use of the system across different devices and networks.

Additionally, PRUs often included multiple categories of problems in a single complaint, reflecting dissatisfaction, inefficacy, and usability issues, such as bulbs disconnecting and requiring constant physical reconfiguration, leading to frustration and loss of trust in the brand. In summary, the PRUs revealed significant problems of behavior, utility, and efficacy in smart bulbs, resulting in frustration and distrust among users.

For the evaluation of the smart water sprinkler, 101 PRUs were manually collected from the Amazon.com website, with ratings of up to three stars, between 2020 and 2022. The analysis revealed that most PRUs are critical and related to behavioral intent, highlighting problems with functionalities such as connection or integration with other devices, aligning with the mobility quality criterion.

The most problematic functionalities identified were connection or integration, irrigation, and automation. Connection issues with Wi-Fi and integration with other devices were common, affecting system efficacy. Irrigation issues were related to physical problems with the device, affecting its efficiency and efficacy. Most PRUs had a negative polarity, highlighting efficacy, efficiency, and utility problems.

Regarding user experience, frustration and lack of confidence predominated. The results indicate unsatisfactory usability and user experience, although it is important to consider that only PRUs with low star ratings were analyzed, which may explain the predominance of negative or neutral sentiments.

The PRUs related to the sprinkler address specific problems related to the quality of use characteristics of ubiquitous systems, highlighting the need for improvements to meet user expectations and needs.

Finally, for the evaluation of the smart lock, 75 PRUs were manually collected from the Amazon.com website, all with ratings of up to three stars, between 2020 and 2022. Most PRUs are critical and related to the behavioral intent of the system.

The main functionality problems were connection and integration with other devices, fingerprint reader, and virtual keyboard. Some comments highlighted the influence of the physical environment on the device's operation. The most frequent usability problems were efficacy, efficiency, and utility. Users expressed dissatisfaction mainly with the efficiency in configuring and using the system.

User experience was predominantly marked by frustration and lack of confidence in the device. In summary, overall usability and user experience were unsatisfactory, failing to meet positive aspects such as satisfaction and motivation.

When evaluating the three devices with the original MALTU, we sought to observe PRUs that possibly related to the quality characteristics of Ubicomp and IoT systems. For example, in this PRU from the smart lock: *"I wouldn't buy again for one simple reason: it jams when it's open, randomly, and the company says it's supposed to do that. So, every once in a while, when you close the door, the bolt fires and BANG, hits the jam. This isn't because it's set up wrong, and I have the sensor installed so it knows when it's closed - but it doesn't care."*, the smart lock fails to perceive environmental information through its sensor, i.e., whether the door is closed or open, thus it can be associated with **sensitivity to context**, just like this PRU from the smart sprinkler: *"This is by far the best water timer I've ever had, with one exception. It can't automatically follow watering restrictions on odd/even days that are common in the US. Unfortunately, being the best doesn't resolve that I have to change the schedule every week or risk breaking the law. For those who don't know, in this system you can only water on odd days if your house number ends in odd and vice versa."*. It's also possible to associate it with **sensitivity to context**, since the system is not able to perceive contextual information from the user's environment, such as the number of their house, the days of the week being odd or even, or even the restrictions caused by a law.

It is also possible to relate some comments to **mobility**, for example in *"The bulb doesn't work well. All others from other brands we have zero problems but this one is already an exchange because the first one just broke and now it keeps presenting a problem. It just doesn't obey the command, neither by the brand's App, nor by the Alexa App, nor by the Alexa device. I do not recommend"*. It is possible to perceive that the bulb does not perform its functions properly as the user moves through different devices, i.e., the bulb's app, the Alexa app, and Alexa itself, or in this PRU from the sprinkler: *"It connects to my home wi-fi quickly, but my husband and I have tried several ways to link it to Google's assistant, but it has failed to do its job. My other home devices have no problems with this."*; it is possible to perceive that the sprinkler does not perform its functions properly as the user tries to link the device to Google's assistant.

6.2 Extended MALTU

This stage was based on [Rusu *et al.*, 2011] in describing the most important attributes collected in the literature to formalize the main concepts that the extension would address in order to cover the domain of the evaluated application. In the case of UbiComp and IoT systems, these attributes correspond to the AMICCaS [Carvalho, 2019], and from this, as MALTU enables other forms of PRU extraction and classification [Mendes, 2015], a new category was proposed that incorporated the definitions of these interaction quality criteria.

The proposed MALTU extension answers RQ1: Is it possible to satisfactorily extend the MALTU model to meet the UbiComp/IoT system quality characteristics (AMICCaS)?

Below is a list of definitions of AMICCaS, along with questions that can help the evaluator decide which problem the PRU is related to:

- **Invisibility**: Hardware resources are visible or attract the user's attention, and system tasks depend on explicit user inputs. Are hardware resources visible or attracting the user's attention? Do the tasks users perform depend on their explicit inputs?
- **Attention**: Technology distracts users from their daily activities; Is the user focusing on the system or on their daily activity(s)?
- **Calmness**: The user feels overwhelmed by the system; Is the system overwhelming the user?
- **Mobility**: The user cannot use the system on different devices or networks; Has system usage been interrupted due to integration or connection problems with other devices or networks?
- **Synchronicity**: Things are not synchronized with each other; Are data and information synchronized across different devices?
- **Context Awareness**: The system fails to perceive contextual information related to the user, the system's environment; Is there contextual information about the user or the environment that the system is not perceiving?
  In Figure 4 it is possible to visually observe how the new definitions were incorporated into the extension.

In addition to these desirable characteristics, it was necessary to make a change in the classification related to the artifact. In the original model, this category refers to which device the user is commenting on in the PRU. In the case of the extension, it is more interesting to know which devices the user is mentioning in the PRUs, since ubiquitous systems are designed to connect with various devices and applications. Furthermore, this modification can help identify where issues of synchronicity and, especially, mobility are occurring. Therefore, the questions following the definitions serve to assist the evaluator in reasoning whether it is indeed possible to extend the MALTU to meet the quality characteristics of UbiComp/IoT systems (AMICCaS).

7 Initial uses of extended MALTU

The extended version of MALTU was used in two works: 1) in an initial validation of the extended MALTU proposal,

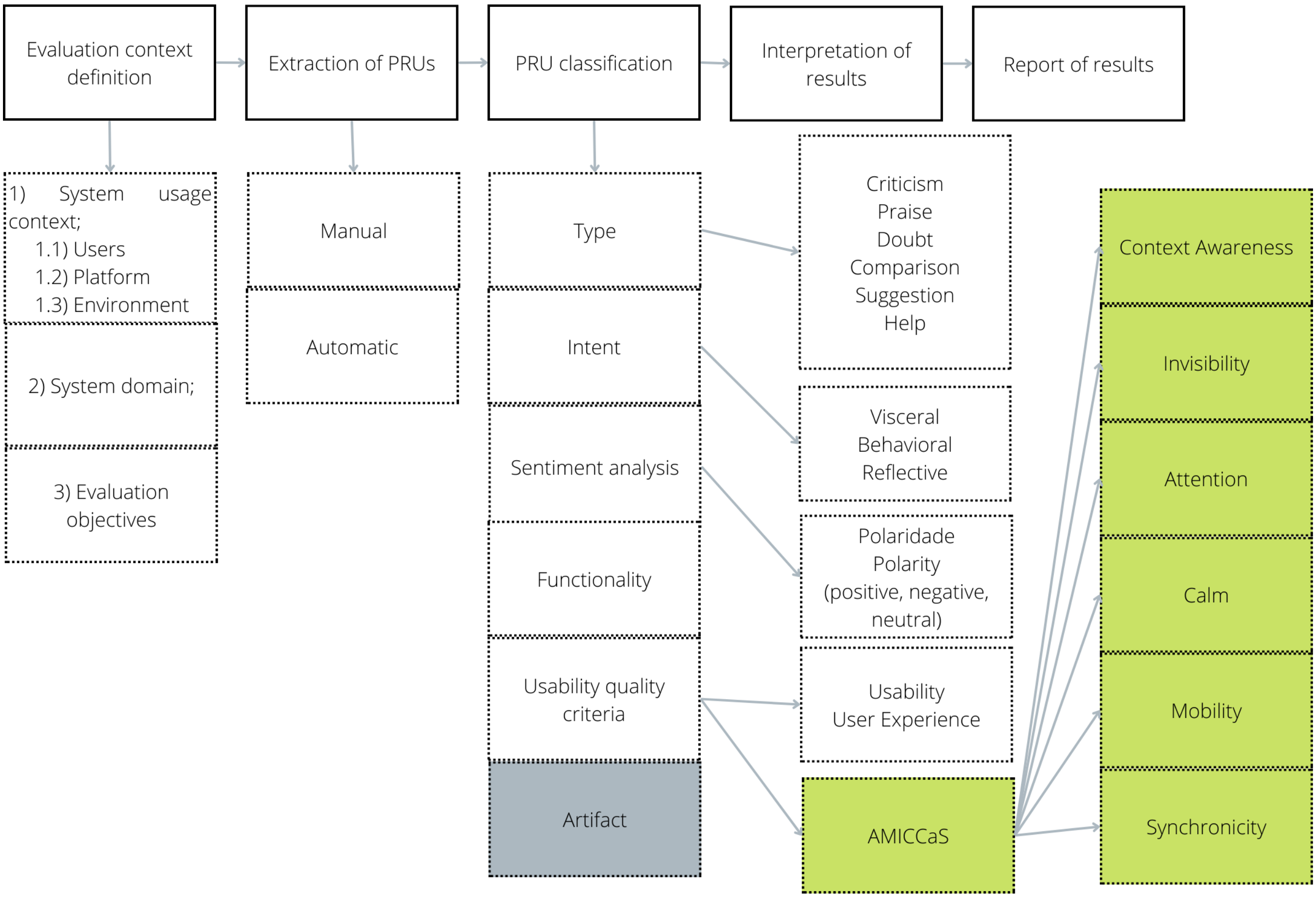

Figure 4. Extended MALTU Schema

which was conducted by the first authors of the extension [omitted for anonymous review]; 2) in an totally independent evaluation [Moreira and Coutinho, 2023]. Both are detailed in the next sections.

This section aims at address the RQ2. How does the proposed extension of MALTU assist HCI professionals in reflecting on the usability characteristics of UbiComp and IoT systems?

7.1 Evaluation 1 (XIAOMI Mi Band 6 smartwatch)

In this first evaluation of the MALTU extension, a device was chosen, following the selection criteria based on Brito [2018]: Having a associated mobile application published on the PlayStore; Having available comments posted by users about their experiences; and the physical device must be available to the evaluators. The evaluation was planned and managed by the first author, but was conducted by three invited evaluators. At this point of the study, we were able to recruit only three evaluators focused on qualitative research, with the objective of a initial validation of the proposed model.

The evaluators were selected for convenience, being advanced semester students in Digital Design who have already taken courses in HCI and HCI Evaluation. All three agreed to the consent form[2] and responded to an initial questionnaire about their profiles[3] that can be visualized in Table 1. The levels (numbers) of experience in the table means: 1: I do not have any experience with the subject; 2: I Know the concepts, but have had no experience with them; 3: I am familiar with the concepts and have had some experience with the subject; 4: I am an expert on the subject and have had several experiences with the area.

The evaluators have over four years of experience in HCI evaluation in desktop systems, using techniques such as interviews, questionnaires, as well as usability testing, focus groups, and heuristic evaluation, which allows them to be described as HCI experts.

Table 1. Evaluators profile

ID	HCI area	Usability evaluation methods	Use of UbiComp systems	MALTU
E1	3	3	3	3
E2	3	3	3	2
E3	3	3	3	2

The selected system was the XIAOMI Mi Band 6 smartwatch (Figure 5), classified as “Smart Wearables”, which features functions such as heart rate monitoring, sleep tracking, blood oxygen monitoring, physical activity tracking with 30 available exercise modes, Bluetooth connection with the mobile app (Zepp Life), among others.

The evaluators’ objective was for each one to individually perform two consecutive evaluations on the same device: the first one with the extended MALTU and the second one was the heuristic evaluation with the Usability Heuristics for

[2] [omitted for anonymous review]
[3] [omitted for anonymous review]

Figure 5. Smartwatches used in the heuristic evaluation

Ubiquitous Systems (HUbis). One reason for choosing this device was the availability of three of them to perform the heuristic evaluation [de Souza Filho *et al*., 2020; Brito, 2018; Rocha, 2017].

There were a total of four online meetings with the evaluators, all conducted by the first author of this paper via Google Meet. These meetings were held sequentially, with short intervals between them to allow for analysis and preparation of subsequent steps. In the first meeting, the steps of the extended MALTU model were explained, with the evaluation context already filled out and the PRUs already extracted, with a Google document available describing the main concepts of the model and classifications for consultation as needed. Each evaluator classified exactly the same PRUs provided in the Google spreadsheet, and none had access to each other's responses. PRU classification lasted 15 days.

In the second meeting, after the conclusion of the first evaluation with the extended MALTU, the results were reported, followed by an explanation of how the heuristic evaluation with the HUbis worked, with the presentation of the basic steps of a heuristic evaluation, such as preparation, learning about the current situation, domain, context, and users [Barbosa and Silva, 2010]. At this stage, three identical smartwatch devices from the same brand were distributed to the evaluators to start the evaluation, with the following tasks to be evaluated:

- T1 - Connection with the application Zepp Life
- T2 - Set alarm
- T3 - Listening to music
- T4 - Training function
- T5 - Sleep monitoring
- T6 - Checking Stress

In the third meeting, after the completion of the heuristic evaluation lasting 6 days, the evaluators met for the results session, which corresponds to a discussion, prioritization of problems, and suggestions for solutions [Rocha, 2017].

Finally, a focus group was conducted with the evaluators to understand their views on the two evaluation methods, guided by the questions bellow. The session was transcribed by the first author, who also carried out a thematic analysis of the excerpts to identify key insights and recurring patterns.

- Q1 – What do you think of the extended MALTU model?
- Q2 – Did you consult the descriptions?
- Q3 – Which category did you find most difficult to classify the comments?
- Q4 – Was there any situation of doubt in the classification?
- Q5 – Did you have any doubt about the AMICCaS classification?
- Q6 – What would you say about the usability of this ubiquitous system based on what you learned from the AMICCaS classification?
- Q7 – What would you say about the smartwatch's usability based on what you learned from the extended MALTU evaluation?
- Q8 – What would you say about the user experience based on what you learned from the extended MALTU evaluation?
- Q9 – What are the main differences you noticed when using the two methods to evaluate the same device?
- Q10 – Were there any interaction problems detected in MALTU that were not addressed by the heuristic evaluation with the HUbis and vice versa?
- Q11 – Is there anything else you would like to comment on about the evaluations?

Regarding the results of the evaluation with the extended MALTU, we can highlight that there were a total of 91 PRUs analyzed by the three evaluators. The occurrences of each AMICCaS category is detailed in in Table 2.

Table 2. Number of PRUs classified in each AMICCaS category

	E1	E2	E3
Context sensitivity	29,67%	19,78%	52,75%
Invisibility	2,20%	0	0
Attention	1,10%	1,10%	0
Calmness	0	0	0
Mobility	0	1,10%	4,4%
Synchronicity	0	16,48%	10,99%
Without classification	49,45%	61,54%	31,87%

The majority of PRUs indicate issues with context sensitivity and synchronization. This suggests that, based on the PRUs analyzed, the system failed to deliver the desired outcome based on the user's contextual information or to synchronize data with the various devices that the bracelet is designed to connect with. An example of a PRU illustrating this scenario is: *"I didn't like that the language isn't in Portuguese and because of this I am having trouble using it; I can't change it."* Since the language is a contextual piece of information and the bracelet was set to a different language, the user was unable to change it, thus also unable to utilize the system's features to support their goals.

In relation to the heuristic evaluation, the violated heuriscts can be seen in Table 3. Considering only the heuristics most related to the specific characteristics of ubiquitous systems (UH11-UH15), the violated ones were UH14 (context sensitivity and adaptive interfaces) and UH15 (sensors and data input). The most impacted ubiquitous characteristics were context sensitivity, adaptation, usability, and data input. This demonstrates issues related to information transmission and updating through communication between devices.

On the connection with the app (T1), E3 reported that the app did not update in relation to the watch's data, experiencing a delay, and added that it was necessary to restart the app to reconnect with the bracelet and update the data. He identified the main functions affected by the delay in updating the data from the bracelet in the app as stress monitoring and heart

Table 3. Heuristics violated in evaluation

	E1	E2	E3
T1	UH5, UH9, **UH14**	**UH10**	**UH15**
T2	UH5, UH9	-	UH1
T3	-	UH9, **UH10**	UH6
T4	UH5, UH9, **UH10**	**UH15**	-
T5	UH5, UH9, **UH14**	-	**UH14**
T6	UH5, UH9	-	**UH14**, **UH15**

rate tracking. Additionally, this evaluator's phone was a XIAOMI Redmi 9, and since it is from the same company, better synchronization was expected. E2 did not see this as a problem, even though he used a phone from a different brand, because, according to him, such data synchronization delays are expected. Moreover, E2 pointed out issues with UH10 because his device was in Mandarin; however, it was possible to synchronize it with his phone using the icons in the interface.

Regarding the alarm feature, E3 noted that there was insufficient feedback on the interface to identify whether the alarm was set or not, as the only visual feedback provided by the bracelet on the interface was a small circle on the alarm screen. To be sure it was set, he had to access the alarm setting to see this visual feedback. E2 also mentioned being somewhat confused about changing the alarm time, but exploring the watch's functions helped him figure out how to make this adjustment.

In task T3, one of the main difficulties was discovering the procedure to authorize playing music through the bracelet. The bracelet requests permission to connect with the phone's music, and its interface explains that app permissions need to be adjusted. However, the instructions are superficial and do not clearly show the path the user must follow to adjust the permissions. One evaluator identified that the bracelet interface incorrectly indicates where the permission authorization path is located. The area needed to authorize music from the phone is hidden among various options in the app, making it more complex and difficult to find, causing the user to seek information online. E3 suggested having a popup menu in the app requesting all necessary permissions for using the phone's data.

In task T4, one of the issues raised was the accuracy of the sensors, as one evaluator reported that the sensors were counting arm movements as steps. The training function is much more precise because it requires activating the GPS; however, if there is a pause of 10 to 20 seconds, the bracelet vibrates, and the training session ends. E2 stated that this duration is too short, as such a pause does not necessarily mean the user has finished the workout; it could just be a brief break to catch one's breath. Despite the issue of accuracy, another evaluator found the information provided by this function useful, such as activity time, distance, route, and calories, especially for those interested in losing weight or engaging in physical activities.

All three evaluators agreed that the sleep monitoring function (T5) is the most comprehensive feature of the bracelet. One commented on the confusion about how the watch identified that he was in REM (Rapid Eye Movement) sleep, a stage of sleep characterized by rapid eye movements. An issue raised by the evaluators was the ergonomics of the bracelet, as wearing it during sleep caused discomfort, itching, and pain in the wrist.

One evaluator reported that the stress monitoring function (T6) only worked on one day because, according to him, his peak stress occurs at noon; however, the watch was not accurate in capturing this data, indicating his stress levels were normal even though he was stressed.

In relation to the focus group results, the descriptions in the statement for Question 2 refer to a document containing the key definitions of the extended MALTU model, such as categories, classification examples, model steps, and primarily explanations of the AMICCaS category. In Question 1, an evaluator noted that a drawback of the model is that user comments often do not contribute relevant information for understanding interaction problems, as they might discuss product delivery or authenticity, rather than system usage. The same evaluator also noted that user reviews (PRUs) sometimes addressed issues concerning both the system and its associated app. Another evaluator suggested filtering out irrelevant comments for a more focused evaluation.

Regarding Question 3, all evaluators agreed that the AMICCaS category was the most difficult to classify. This difficulty stemmed from the fact that, although the evaluators were familiar with using ubiquitous systems in their daily lives, they were not acquainted with the specific quality characteristics required to assess such systems. This gap made it challenging to interpret and apply the technical definitions when categorizing user comments.

In response to Question 4, there were indeed situations in which the evaluators had doubts during classification. Some PRUs contained mixed content, such as a comment beginning with praise and immediately followed by criticism. This overlap led to uncertainty about whether a single comment should be classified under one or multiple types, making the classification process more complex and interpretative.

Question 5 confirmed that the evaluators experienced doubts specifically when classifying content under the AMICCaS category. Despite being frequent users of ubiquitous systems, they found this classification to be more technical and nuanced, requiring familiarity with system-specific quality features that were not part of their everyday usage knowledge.

All three evaluators referred to the descriptions, with one highlighting the difficulty in classifying comments under the AMICCaS category, as these comments typically address system-specific issues, which were not frequently discussed in the PRUs. An example shared during a focus group involved a PRU where a user complained about the device's language being in Mandarin; this evaluator struggled to categorize this issue but ultimately associated it with context sensitivity, which seemed the most relevant to the problem described.

Another evaluator discussed the challenge in responding to Question 6, as the PRUs did not directly pertain to the system itself. However, they noted from the AMICCaS category that the device exhibited synchronization problems, particularly during training activities where running distances were not accurately logged, and difficulties in connecting the phone with the bracelet were observed. Additionally, it was identified that users did not check whether the product was the global version, which may have caused the language issue.

In Question 7, the evaluators found that the watch was effective, with a matching interface between the bracelet and the app, and the sensors were operational though not highly accurate. Moreover, an evaluator reported learning and memorability issues during their first use of the smartwatch. Some of the problems included a non-functional bracelet, negatively impacting user satisfaction, and inefficacy in certain functionalities like a static blood pressure monitor and a non-responsive touchscreen.

Question 8 highlighted frustration as the most commonly identified sentiment in the PRUs. However, one evaluator recalled a PRU where a user received a counterfeit bracelet that still functioned, which to the user was acceptable, highlighting the theme of trust. Additionally, this evaluator pointed out that in PRUs containing praise, it was not possible to ascertain how long the user had been using the watch. Other issues included inaccurate distance tracking by the running sensors, leading to user frustration due to unmet expectations.

In Question 9, one evaluator noted that the heuristic evaluation provided a clearer and more direct observation of the bracelet's real issues, stating that PRUs did not fully capture the conditions under which users faced problems, unlike the heuristic evaluation with the HUbis, which offered a comprehensive view of both the device's positives and negatives. The broad nature of PRUs made classification laborious, necessitating the filtration of relevant information for focused evaluation. While another evaluator found the extended MALTU model comprehensive, it aided the heuristic evaluation by providing a broad overview of the product aspects being evaluated. However, they noted that heuristic evaluations could be biased based on the evaluator's profile, as issues like synchronization and input sensors were not severe for an evaluator not focused on physical activities, unlike another evaluator who considered these issues critical for users actively using the training data to enhance their physical conditions.

The primary difference between the two evaluation methods discussed in Question 9 is their scope, with heuristic evaluations using HUbis being better suited for identifying system issues. However, the extended MALTU model addressed issues not observable in heuristic evaluations, such as device language settings, touchscreen functionality, synchronization problems, and even delivery issues indirectly affecting user satisfaction. Conversely, heuristic evaluations identified issues such as step counting based solely on arm movement, which were not covered in PRUs, and problems related to music control functions, which were more problematic and complex to execute than described in PRUs.

As for Question 10, the evaluations revealed that certain interaction issues were uniquely captured by each method. For example, the extended MALTU evaluation helped identify a problem related to context sensitivity, specifically concerning the system's language being set to Mandarin, which was not apparent in the product description or expected by the users. On the other hand, the heuristic evaluation with the HUbis enabled real-time detection of synchronization issues, as the evaluators used the device and observed problems such as data not updating properly between the bracelet and the app.

In response to Question 11, the evaluators noted that their experience with the extended MALTU model could have been improved if they had received a brief introductory session or explanation specifically focused on the AMICCaS category. This would have better prepared them to understand and apply the quality characteristics associated with ubiquitous systems, reducing doubts and enhancing the consistency of the classification process.

7.2 Evaluation 2 (G-TECH App and FreeStyle Libre)

This study revolves around two UbiComp systems designed to facilitate blood glucose monitoring: the G-TECH APP and the FreeStyle Libre. In this research endeavor, the context of use, system domain, and evaluation objectives of both systems are explored comprehensively to shed light on their effectiveness and user experience.

The G-TECH APP (Figure 6) caters to individuals seeking a convenient method for monitoring blood glucose levels. It integrates seamlessly with a Bluetooth glucometer and test strips, enabling users to perform readings on-the-go. Its target audience primarily comprises people with type 1 diabetes who require a reliable and efficient solution for managing their condition. The system operates within various settings, including restaurants, hotels, and residential environments, utilizing smartphones as the primary platform for interaction.

Figure 6. Home screen of G-TECH App

In contrast, FreeStyle Libre (Figure 7) offers a more advanced approach to blood glucose monitoring, employing a subcutaneous sensor for continuous readings. This sensor, coupled with smartphone integration, provides users with real-time insights into their glucose levels, facilitating better management of their condition. Like the G-TECH APP, FreeStyle Libre targets individuals with type 1 diabetes, offering enhanced functionalities such as continuous readings and third-party access to data via the app.

The evaluation objectives for both systems are centered around assessing the user experience of individuals with type 1 diabetes and identifying any potential issues or areas for improvement. By conducting a thorough evaluation of these systems, this study aimed to contribute valuable insights into their usability, effectiveness, and overall impact on diabetes management.

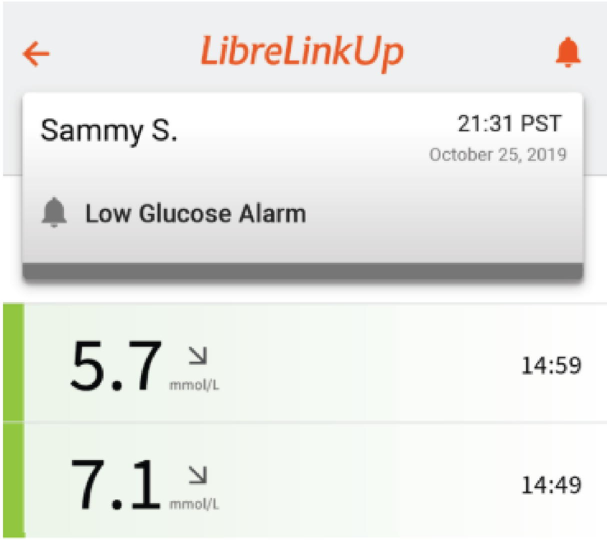

Figure 7. Home screen of FreeStyle Libre

After defining the context, PRUs were extracted manually from social systems such as Amazon, App Store, and Play Store. A total of 114 PRUs related to the G-TECH App were evaluated, with 5 being excluded from the final assessment, along with 23 PRUs related to the FreeStyle Libre. This discrepancy in numbers is attributed to the varying availability of reviews for each application. The PRUs were collected between 2020 and 2023 from the Play Store, App Store, and Amazon platforms, specifically focusing on reviews rated with 3 stars or fewer, as they typically highlight issues within the application or system.

All PRUs were individually evaluated and classified, with revisions conducted on separate dates for each assessment. The evaluation period spanned from November 20, 2022, to June 18, 2023. To maintain ethical standards, the privacy of reviewers was respected by not collecting their names; only their comments and posting dates were recorded for differentiation. No interviews were conducted for this research, thus eliminating the need for experiments involving human subjects.

Regarding the results, the majority of PRUs were extracted from the Play Store (78.9%), followed by Amazon (11.9%) and the App Store (9.1%), spanning the years 2020 to 2023. The results revealed a significant number of PRUs related to synchronization (54.1%) in the G-TECH App, reading recording (15.6%), and record exportation (9.2%). For the FreeStyle Libre, 30.4% of PRUs were also related to synchronization, 13% concerning notifications, and 13% addressing discrepant records.

Furthermore, it was used the AMICCaS classification provided in the extended MALTU. In the research, the author identified that 63.3% of the extracted PRUs from the G-TECH app pertained to Synchronicity, while 36.7% were not identified with the category classifications. In contrast, for the FreeStyle Libre, the author found that 34.8% of the extracted PRUs addressed AMICCaS Synchronicity, and another 34.8% were related to Context Sensitivity. Additionally, the author also identified that 13% of PRUs were about Mobility.

8 Discussion and final remarks

The Evaluation 1 of the smartwatch revealed significant challenges in classifying the AMICCaS within the evaluated systems, highlighting a disconnection between the evaluators' practical knowledge and the theoretical depth required by the concepts of usability quality criteria in UbiComp and IoT systems in the extended MALTU model. Most evaluators, mostly familiar only with the everyday use of ubiquitous computing, struggled to classify the PRUs in the AMICCaS classification in practical evaluations. This suggests a significant barrier in theoretical domain that can be mitigated through specific training or revision in the definitions used to ensure greater clarity and applicability.

Both in Evaluation 1 and 2, it was possible to observe a concentration on specific AMICCaS; it is even possible to observe that some of the AMICCaS did not have evidence. In Evaluation 1, for example, none of the evaluators could identify PRUs addressing Calmness; however, in Evaluation 2, there were no PRUs identifying Attention or Invisibility, but there was 8.7% identifying Calmness.

The analysis of invisibility in the evaluated systems pointed to a dual perspective: physical and usage. While the former refers to the non-obtrusiveness of the devices themselves, the latter addresses the minimization of user interaction with the system. This distinction is crucial to understand how UbiComp and IoT systems can be optimized to become more integrative and less intrusive in everyday environments. There is a need for restructuring the definitions of invisibility, as each evaluator may have one of the two views or even a different view of what invisibility of UbiComp and IoT Systems might mean. To address this, it is necessary to identify if it is possible to find PRUs that address both one and the other side.

The results of Evaluation 1 also point to the effectiveness of heuristic evaluations in identifying specific usability problems that PRUs could not fully capture. The feedback from the evaluators indicates that greater depth of analysis could be achieved with more evaluation time, suggesting that time constraints may compromise the quality of insights obtained. This raises a question about the ideal balance between depth and efficiency in the heuristic evaluation process in ubiquitous systems contexts. In this context, the use of the MALTU model, for its speed, presents itself as a viable possibility in the face of time limitations.

However, in addition to the time constraints, other limitations may have influenced the results of Evaluation 1. The study design, based on prior exposure of the evaluators to the product and tasks, may have introduced bias in the identification of problems. Furthermore, the evaluators' prior knowledge—limited mainly to their everyday experience with technology—may have influenced their interpretation and classification of PRUs, particularly in more technical categories such as AMICCaS. These factors highlight the importance of carefully considering the expertise of participants and the setup of evaluation sessions when applying and validating extended models.

The evaluation of the G-TECH APP and FreeStyle Libre systems in Moreira and Coutinho [2023] is closely aligned with the previous discussion on the challenges and insights derived from the extended MALTU model. Both systems, designed to enhance the daily management of type 1 diabetes, incorporate principles that resonate with the themes of invisibility and usability discussed earlier.

The AMICCaS classification, which was highlighted as complex in previous findings, plays a critical role in understanding user feedback in the context of these ubiquitous health technologies. The substantial focus on Synchronicity in the G-TECH APP (63.3%) and a significant portion in FreeStyle Libre (34.8%) underscores the paramount importance of seamless operation within these systems. This aligns with earlier discussions that improved definitions and clarity in the application of AMICCaS could significantly enhance how usability issues are diagnosed and addressed.

Both systems strive for a form of invisibility – the G-TECH APP through its integration with everyday mobile devices and the FreeStyle Libre through its subcutaneous sensor that minimizes interaction. This echoes the refined definitions of invisibility from the extended MALTU model, emphasizing the need for systems to be unobtrusive and minimally invasive in the user's environment and interaction. The goal is for the technology to blend into the user's lifestyle, reducing the cognitive and physical burden typically associated with chronic disease management.

Another discussion point that we can raise is that one of the evaluators from Evaluation 1 commented that the extended MALTU helped him to get an overall picture of the possible problems that could be found in the heuristic evaluation with Hubis. The MALTU can serve to give an overview of the problems that real users experience when using the system.

In light of the results, it is important to revisit the research questions. Regarding RQ1 — whether it is possible to satisfactorily extend the MALTU model to meet the UbiComp/IoT system quality characteristics (AMICCaS) — the findings indicate that the extension is feasible and useful. However, its effectiveness is conditioned by the evaluators' understanding of the model, particularly of the more technical categories. This highlights the need for clearer guidelines or introductory training to facilitate the application of the extended MALTU.

As for RQ2 — how the proposed extension helps HCI professionals reflect on quality use characteristics — the results suggest that the model provides a valuable framework for identifying issues that may not be readily apparent through traditional heuristic evaluation. For instance, characteristics like Calmness or Invisibility emerged more clearly when evaluators engaged with the extended categories, allowing them to reflect on aspects of interaction beyond functionality. Nonetheless, the model's impact depends significantly on the evaluator's profile and the contextual richness of the data (e.g., PRUs).

Given the observations and results of the two evaluations, several directions for future research can be identified. Firstly, extending the MALTU model to cover other usability quality criteria may enrich the applicability of the model in different technological contexts. Additionally, creating a specific guide for developing extensions of the MALTU could facilitate its adaptation and use by other researchers and practitioners. Lastly, exploring the MALTU in other technological domains, such as embedded or expert systems, could validate the robustness and flexibility of the model in various technological applications.

Declarations

Authors' Contributions

Carlos Feitosa and Francisco Catunda Neto performed the experiments reported in this article. Both are the main contributors and writers of this manuscript. Ingrid Monteiro supervised the work. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

The authors acknowledge the voluntary participation of the individuals involved in the studies reported in this article.

Availability of data and materials

The datasets generated during the current study will be made available upon request.

References

Andrade, R. M. C., Carvalho, R. M., de Araújo, I. L., Oliveira, K. M., and Maia, M. E. F. (2017). What changes from ubiquitous computing to internet of things in interaction evaluation? page 3–21. DOI: 10.1007/978-3-319-58697-7_1.

Asghari, P., Rahmani, A. M., and Javadi, H. H. S. (2019). Internet of things applications: A systematic review. *Computer Networks*, 148:241–261. DOI: 10.1016/j.comnet.2018.12.008.

Barbosa, S. and Silva, B. (2010). *Interação humano-computador*. Elsevier Brasil. Book.

Brito, M. R. F. (2018). Explorando o modelo maltu na avaliação de usabilidade e experiência de uso em um dispositivo ubíquo. Available at:`http://repositorio.ufc.br/handle/riufc/39486`.

Carvalho, R. M. (2017). Dealing with conflicts between non-functional requirements of ubicomp and iot applications. In *2017 IEEE 25th International Requirements Engineering Conference (RE)*, pages 544–549. DOI: 10.1109/RE.2017.51.

Carvalho, R. M. (2019). Correlate & lead: process and catalog of non-functional requirements correlations in ubicomp and iot systems. Available at:`http://repositorio.ufc.br/handle/riufc/47680`.

Carvalho, R. M., de Castro Andrade, R. M., and de Oliveira, K. M. (2018). Aquarium-a suite of software measures for hci quality evaluation of ubiquitous mobile applications. *Journal of Systems and Software*, 136:101–136. DOI: 10.1016/j.jss.2017.11.022.

Carvalho, R. M., de Castro Andrade, R. M., de Oliveira, K. M., de Sousa Santos, I., and Bezerra, C. I. M. (2017).

Quality characteristics and measures for human–computer interaction evaluation in ubiquitous systems. *Software Quality Journal*, 25:743–795. DOI: 10.1007/s11219-016-9320-z.

De Souza, C. S. (2005). *The semiotic engineering of human-computer interaction*. MIT press. DOI: 10.7551/mit-press/6175.001.0001.

de Souza, L. G., Barbosa, S. D. J., and Conte, T. (2015). Evaluating a molic extension for collaborative systems design. In *Proceedings of the 14th Brazilian Symposium on Human Factors in Computing Systems*, IHC '15, New York, NY, USA. Association for Computing Machinery. DOI: 10.1145/3148456.3148471.

de Souza Filho, J. C., Brito, M. R. F., Mendonça, A. R. R., Martins, M. V., and Sampaio, A. L. (2018). Hidrate spark: Avaliando um sistema ubíquo para motivar a ingestão de Água. In *Anais Estendidos do XVII Simpósio Brasileiro sobre Fatores Humanos em Sistemas Computacionais*, Porto Alegre, RS, Brasil. SBC. DOI: 10.5753/ihc.2018.4204.

de Souza Filho, J. C., Brito, M. R. F., and Sampaio, A. L. (2020). Comparing heuristic evaluation and maltu model in interaction evaluation of ubiquitous systems. In *Proceedings of the 19th Brazilian Symposium on Human Factors in Computing Systems*, pages 1–10. DOI: 10.1145/3424953.3426639.

de Souza Filho, J. C., Sampaio, A. L., and Feitosa, E. L. (2022). Hidden things and implicit interaction: Looking back on how invisibility evolved as a quality-in-use attribute. In *Proceedings of the 21st Brazilian Symposium on Human Factors in Computing Systems*, IHC '22. Association for Computing Machinery. DOI: 10.1145/3554364.3559136.

Diniz, L. D. N., Filho, J. C. D. S., and Carvalho, R. M. (2022). Can user reviews indicate usability heuristic issues? *Conference on Human Factors in Computing Systems - Proceedings*. DOI: 10.1145/3491101.3519659.

Freitas, L. M., da Silva, T. H. O., and Mendes, M. S. (2016). Evaluation of spotify: An evaluation textual experience using the maltu methodology. In *Proceedings of the 15th Brazilian Symposium on Human Factors in Computing Systems*, IHC '16, New York, NY, USA. Association for Computing Machinery. DOI: 10.1145/3033701.3033752.

Karvonen, H. and Kujala, T. (2014). Designing and evaluating ubicomp characteristics of intelligent in-car systems. *Advances in Ergonomics In Design, Usability & Special Populations: Part I*, 16:27. DOI: 10.54941/ahfe100765.

Korhonen, H., Arrasvuori, J., and Väänänen-Vainio-Mattila, K. (2010). Let users tell the story: evaluating user experience with experience reports. In *CHI'10 Extended Abstracts on Human Factors in Computing Systems*, pages 4051–4056. DOI: 10.1145/1753846.1754101.

Kourouthanassis, P. E., Giaglis, G. M., and Karaiskos, D. C. (2008). Delineating the degree of'pervasiveness' in pervasive information systems: an assessment framework and design implications. In *2008 Panhellenic Conference on Informatics*, pages 251–255. IEEE. DOI: 10.1109/pci.2008.28.

Melo, B., Andrade, R., and Darin, T. (2022). Longitudinal user experience studies in the iot domain: a brief panorama and challenges to overcome. Association for Computing Machinery. DOI: 10.1145/3554364.3559135.

Mendes, M. S. (2015). Maltu–um modelo para avaliação da interação em sistemas sociais a partir da linguagem textual do usuário. Available at:`http://repositorio.ufc.br/handle/riufc/19820`.

Mendes, M. S. and Furtado, E. (2018). An experience of textual evaluation using the maltu methodology. In *Social Computing and Social Media. Technologies and Analytics: 10th International Conference, SCSM 2018, Held as Part of HCI International 2018, Las Vegas, NV, USA, July 15-20, 2018, Proceedings, Part II 10*, pages 236–246. Springer. DOI: 10.1007/978-3-319-91485-5$_1$8.

Metsis, V., Le, Z., Lei, Y., and Makedon, F. (2008). Towards an evaluation framework for assistive environments. In *Proceedings of the 1st international conference on PErvasive Technologies Related to Assistive Environments*, pages 1–8. DOI: 10.1145/1389586.1389601.

Moreira, F. and Coutinho, E. (2023). Evaluating the user experience of type 1 diabetes control applications. In *Proceedings of the XXII Brazilian Symposium on Human Factors in Computing Systems*, pages 1–11. DOI: 10.1145/3638067.3638134.

Nielsen, J. and Molich, R. (1990). rochas. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '90, page 249–256, New York, NY, USA. Association for Computing Machinery. DOI: 10.1145/97243.97281.

Ntoa, S., Margetis, G., Antona, M., and Stephanidis, C. (2021). User experience evaluation in intelligent environments: A comprehensive framework. *Technologies*, 9(2):41. DOI: 10.3390/technologies9020041.

Oliveira, E. R., Luz, L. C., and Prates, R. O. (2008). Aplicação semi-estruturada do método de inspeção semiótica: estudo de caso para o domínio educacional. In *Proceedings of the VIII Brazilian Symposium on Human Factors in Computing Systems*, pages 50–59. Available at:`https://dl.acm.org/doi/10.5555/1497470.1497477?__cf_chl_f_tk=O9uMcWlg9FCHbGkSpiPT1qu7fHkWPSIXquWOrf9vzOU-1783447540-1.0.1.1-cV5Y4lSl9qTILL.HUPpIDyAsEIe6ny1Ak1Q2nhikhm0`.

Olsson, T. and Salo, M. (2012). Narratives of satisfying and unsatisfying experiences of current mobile augmented reality applications. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 2779–2788. DOI: 10.1145/2207676.2208677.

Quiñones, D. and Rusu, C. (2017). How to develop usability heuristics: A systematic literature review. *Computer Standards & Interfaces*, 53:89–122. DOI: 10.1016/j.csi.2017.03.009.

Rocha, L. C. (2017). Hubis: heurísticas de usabilidade para avaliar sistemas ubíquos. Available at:`http://repositorio.ufc.br/handle/riufc/30526`.

Rusu, C., Roncagliolo, S., Rusu, V., and Collazos, C. (2011). A methodology to establish usability heuristics.(2011). Available at:`https://www.researchgate.net/publication/229040164_A_Methodology_to_establish_usability_heuristics`.

Santiago, M. T. and Marques, A. B. (2022). Are user reviews useful for identifying accessibility issues that autistic users face? an exploratory study. *ACM International Conference Proceeding Series*. DOI: 10.1145/3554364.3559114.

Santiago, M. T. and Marques, A. B. (2023). Exploring user reviews to identify accessibility problems in applications for autistic users. *Journal on Interactive Systems*, 14:317–330. DOI: 10.5753/jis.2023.3238.

Santos, P. S. H. D., Oliveira, A. D. A., Jesus, T. B. N. D., Aljedaani, W., and Eler, M. M. (2024). Evolution may come with a price: analyzing user reviews to understand the impact of updates on mobile apps accessibility. Association for Computing Machinery. DOI: 10.1145/3638067.3638081.

Santos, R. M. (2014). Características e medidas de software para avaliação da qualidade da interação humano-computador em sistemas ubíquos. Available at:`https://repositorio.ufc.br/bitstream/riufc/21446/1/2014_dis_rmsantos.pdf`.

Siewerdt, F. L., Carvalho, R. M., and Andrade, R. M. C. (2016). Recommendations for usability testing in ubiquitous applications. In *Proceedings of the 15th Brazilian Symposium on Human Factors in Computing Systems*, IHC '16, New York, NY, USA. Association for Computing Machinery. DOI: 10.1145/3033701.3033749.

Tuch, A. N., Trusell, R., and Hornbæk, K. (2013). Analyzing users' narratives to understand experience with interactive products. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 2079–2088. DOI: 10.1145/2470654.2481285.